DOCUMENT RESUME

ED 114 411 TM 004 902

AUTHOR Gillmore, Gerald M.
TITLE The Relationship Between Graduating Senior Nominations of Valuable and Non-valuable Courses and End-of-Course Student Ratings.
INSTITUTION Washington Univ., Seattle. Educational Assessment Center.
REPORT NO FAC-P-271b
PUB DATE Jul 75
NOTE 12p.

EDRS PRICE MF-$0.76 HC-$1.58 Plus Postage
DESCRIPTORS *College Students; Comparative Analysis; *Course Evaluation; Effective Teaching; Higher Education; Participant Satisfaction; Questionnaires; Seniors; Student Evaluation; Student Opinion; Summative Evaluation; *Teacher Rating; *Time Perspective; *Validity
IDENTIFIERS *Washington University

ABSTRACT

The College of Arts and Sciences Senior Survey asked students to nominate their most and least valuable courses during their undergraduate careers. The end-of-course student ratings were compared between 40 courses rated as valuable and 16 courses rated as nonvaluable. All differences were statistically significant, with valuable courses getting more favorable ratings on all items. Items which most strongly discriminated between the two groups tended to be those addressing broad educational outcomes, while items showing least discrimination dealt with the mechanics of good teaching. (Author)

****
* Documents acquired by ERIC include many informal unpublished *
* materials not available from other sources. ERIC makes every effort *
* to obtain the best copy available. Nevertheless, items of marginal *
* reproducibility are often encountered and this affects the quality *
* of the microfiche and hardcopy reproductions ERIC makes available *
* via the ERIC Document Reproduction Service (EDRS). EDRS is not *
* responsible for the quality of the original document. Reproductions *
* supplied by EDRS are the best that can be made from the original. *
****

ED114411

U S DEPARTMENT OF HEALTH
EDUCATION & WELFARE
NATIONAL INSTITUTE OF
EDUCATION
THIS DOCUMENT HAS BEEN REPRODUCED EXACTLY AS RECEIVED FROM THE PERSON OR ORGANIZATION ORIGINATING IT. POINTS OF VIEW OR OPINIONS STATED DO NOT NECESSARILY REPRESENT OFFICIAL NATIONAL INSTITUTE OF EDUCATION POSITION OR POLICY

ED114411

Educational Assessment Center

University of Washington

July 1975

76-2

The Relationship between Graduating Senior Nominations of Valuable and Non-valuable Courses and End-of-Course Student Ratings

Gerald M. Gillmore

The College of Arts and Sciences Senior Survey asked students to nominate their most and least valuable courses during their undergraduate careers. The end-of-course Student Ratings were compared between forty courses rated as valuable and sixteen courses rated as non-valuable. All differences were statistically significant, with valuable courses getting more favorable ratings on all items. Items which most strongly discriminated between the two groups tended to be those addressing broad educational outcomes, while items showing least discrimination dealt with the mechanics of good teaching.

Educational Assessment Center Project: 271b

The Relationship between Graduating Senior Nominations of Valuable and Non-valuable Courses and End-of-Course Student Ratings

Gerald M. Gillmore

The use of student ratings of instructional effectiveness has shown a marked increase in higher education. However, critics and proponents alike are (or ought to be) wary of an over-emphasis upon this single source of evaluative information at the exclusion of others. A statement "On the Techniques of Teacher Evaluation" issued by the University of Washington Faculty Senate Committee on the Evaluation and Improvement of Teaching, contained the following: "As important as student ratings are, however, they are simply part of the picture and no single technique can adequately measure a person's teaching contribution."

Reasons for the seeming over-reliance on student ratings is probably two-fold: they are relatively easy to collect, and they are psychometrically reliable. Other methods require greater expenditure of valuable resources, such as faculty time, to obtain systematic and reliable information.

There are two approaches to this problem, with the approach chosen having implications for how one views the validity of student ratings data. If one views student ratings as only one source of data, and other sources are to be pursued with diligence, then the validity of student ratings largely comes down to a question of "Is the device collecting information which is an accurate appraisal of student opinion of the value of the course at its end?" One could test this validity by coordinating student rating results with results obtained by concurrently administered alternative evaluational techniques. Correlations between end-of-course student ratings and measures from other sources or points in time would be interesting, especially in respect to learning about the concept of teaching effectiveness, but would have little to say about the validity of student ratings per se.

The alternative approach is to view student ratings not only as a valid measure as defined above but also as a substitute for additional measures. This approach necessarily broadens the validity question

considerably, because now one is talking about the validity of student ratings as a measure of teaching effectiveness and not just as a measure of student opinion about teaching effectiveness. Within this approach, correlations between student ratings and other measures are direct indicators of the concurrent validity of the method.

Depending upon the approach one wishes to adopt, research to be presented here is a study of the relationship between two measures of teaching effectiveness or a study of the validity of student ratings. This particular study concerns the point in time of the student evaluation. One common criticism of end-of-course student ratings is that students do not have the necessary perspective to make an accurate assessment of the value of a course. Anyone attending faculty discussions of student ratings is familiar with the mythical professor whose courses are (were) rated and maligned by the students while enrolled, but dearly loved and respected by these same students upon entry into professional life. The present study compares the opinion of students at the end of their senior year with opinions of students at the end of the course. Presumably, at the end of the senior year, students can look back over their course work with some greater perspective than they can have at the end of each course taken. The absence of a relationship between how valuable a course is, viewed by graduating seniors and the student ratings that same course received, would either highlight the importance of systematically collecting information from graduating seniors for evaluating instructors and courses or throw the validity of student ratings into question.

Method

The Instruments

The Senior Survey. In June of 1974, questionnaires were mailed to all baccalaureate degree candidates within the College of Arts and Sciences at the University of Washington. (For a complete description of the instrument and results, see de Wolf, Note 1). Contained in the questionnaire were the following three requests:

1. Please name three courses and instructors within your major which now seem to have been most valuable in your education at the UW.

2. Please name three courses and instructors outside your major which now seem to have been most valuable in your education at the UW.

3. Please name three courses and instructors which now seem to have been least valuable to your education at the UW.

Responses to these three requests were the only data used from the Senior Survey in this study.

The University of Washington Survey of Student Opinion of Teaching. From 1968 until 1974, the standard form for collection of student ratings data at the University of Washington contained 24 items, only the first 15 of which were used for this study. These items are found in Table 1. Each item employs a five position response scale, with 1 being assigned to the most favorable position and 5 to the least.

Subjects

The Senior Survey was sent to 1,845 students, and returned by 898 or 48 percent of the population. Students completing end-of-class student ratings were those enrolled in the specific class rated. Some subjects completing the Senior Survey may have also been a part of the subjects who completed student ratings in some cases. However, the anonymity of student ratings precludes any determination of this overlap, however it is probably negligible.

Selection of Classes

The unit of analysis for this study was classes, not students. In the Senior Survey a total of 2,641 course-instructor combinations were mentioned one or more times (about half of these were mentioned only once). Courses mentioned without a specific instructor, and instructors mentioned without a specific course were eliminated from consideration. A course-instructor combination, which will be henceforth referred to as simply course, could be nominated by a graduating senior under any one of the three requests: most valuable within major, most valuable outside major, and least valuable. Thus, for each course, a resultant was calculated which summed number of nominations in the first two categories, and subtracted the number in the third. A positive resultant is indicative of

a "valuable" class, a negative resultant is indicative of a "non-valuable" class. The range of the resultants was +51 to -28.

From these data, two groups of courses were formed. Valuable courses were defined as those having a resultant of +6 or greater. Non-valuable courses were defined as those having a resultant of -3 or less. The former minimum was chosen to be twice as great in absolute value as the latter because of the ratio of two positive questions to one negative question. This procedure yielded 64 valuable courses and 30 non-valuable courses.

Next, files were checked to see which courses had been rated by students using the standard University of Washington Student Rating form any time during the years 1968 to 1974. Of the 64 valuable courses, 40 had been rated. Of the 30 non-valuable courses, 16 had been rated. The difference in the proportion of courses found was not significant ($\chi^2 = .74$, df = 1). In those cases where the same course had been rated for more than one offering, one particular section was chosen randomly. Thus, the final sample consisted of 40 valuable courses, and 16 non-valuable courses.

Method of Comparison

The two types of classes were compared on each of the 15 student rating items by use of *t* tests. Also computed for each item was ω^2 (Hays, 1963, p. 327), which is an index of the strength of relationship. It is indicative of the proportion of the variance of the dependent variable which is attributable to the independent variable.

The reader should be cautioned that even though each of the 15 *t* tests are independently computed, the 15 items of the student ratings form are positively intercorrelated. Thus, results should not be interpreted as if there are 15 statistically independent dependent variables.

Results and Discussion

Class Size

The average class size was 82.9 students for the valuable courses, and 66.4 students for the non-valuable courses. This difference was non-significant ($t = .389$), probably in the most part due to the relatively large standard deviations (67.3 and 55.3 respectively). These large average class sizes are no doubt an artifact of the selection method, larger

classes are apt to get more nominations by sheer force of numbers. However, this result does show that a course does not require a small enrollment to be considered valuable. It also suggests that valuable and non-valuable courses are not differentiated by class size, e.g., large classes are not of value, while small classes are.

Student Ratings Items

The results of the student rating item comparisons are found in Table 1. The items have been arranged in order of the magnitude of *t* value and ω^2.

As can be seen in Table 1, the *t* values for all items were highly significant. The means show that the valuable group of classes were given a more favorable average rating in every case.

The ω^2's ranged from .46 to .17, illustrating reasonably strong relationships. To give an alternative indication of the magnitude of the relationships, the frequency distribution of course means within the two groups for item 9, the item which exhibited the strongest relationship, is found in Table 2. The relatively small amount of overlap is readily apparent.

It might be well to remind the reader at this point that these comparisons are not between the graduating seniors' nominations of valuable and non-valuable courses with their ratings of the course at the time in which they were enrolled. The latter data are based on the ratings of a specific course offering and may contain a few of the sample of graduating seniors, but would almost have to contain mostly students not within the sample. In fact, no attempt was made to find the particular course offering in which the seniors were enrolled--if indeed all who nominated a particular course were enrolled in the same offering of it, e.g., Fall Quarter, 1972, as opposed to Fall Quarter, 1973. This is not considered a weakness of this study, however. The uncontrolled variance resulting from choosing a particular course offering would necessarily add to the error variance of the *t* tests and ω^2's and reduce the magnitude of the *t* values and ω^2's by an unknown amount. For example, suppose a teacher offered the same course twice, and one offering was superior to the second. The superior offering would be more apt to be mentioned as a valuable

Table 1

Comparison of Valuable and Non-valuable Courses on 15 Student Rating Items

Item Number	Item	Mean: Valuable courses (N=40)	Mean: Non-valuable courses (N=16)	t-value*	ω^2
9.	Gave me new viewpoints or appreciations	1.79	2.39	-6.94	.46
12.	Material enthusiastically presented in teaching sessions	1.56	2.22	-6.37	.41
4.	Helped broaden my interests	1.85	2.52	-6.05	.39
7.	Motivated me to do my best	2.24	2.93	-5.98	.38
1.	Abstract ideas and theories were clearly interpreted	1.98	2.58	-5.94	.38
2.	Takes an active, personal interest in the class	1.65	2.24	-5.57	.35
3.	My skills in thinking were increased	2.10	2.70	-5.51	.34
5.	Stressed important material	1.33	2.37	-5.37	.33
15.	Integration of material into coherent whole was	2.06	2.63	-5.17	.31
8.	Inspired class confidence in instructor's knowledge of subject	1.61	2.18	-5.14	.31
10.	Clear and understandable in explanations	1.98	2.54	-4.63	.27
11.	Teaching sessions gave views and info readings did not contain	1.86	2.35	-4.37	.24
6.	Made good use of examples and illustrations	1.70	2.18	-4.21	.23
13.	Material presented in a well-organized fashion	1.87	2.38	-3.68	.18
14.	Helpful to individual students	2.07	2.56	-3.51	.17

*$p < .001$ for all items.

Table 2

Frequency Distribution for Item 9 within Groups

Mean	Frequency	
	Valuable courses (N=40)	Non-valuable courses (N=16)
- 139	4	
140 - 159	6	
160 - 179	11	1
180 - 199	8	
200 - 219	8	3
220 - 239	3	5
240 - 259		2
260 - 279		4
280 - 299		0
300 -		1

course but no more apt to be chosen as the course within the sample analyzed. Therefore, highly significant and strong relationships formed might be considered all the more impressive. Certainly a strong relationship between the alternative methods is in strong evidence.

The items in Table 1 were ordered by the size of the t values and, equivalently, ω^2's in Table 1. This was done in order to speculate from the data about what is important in producing a class which will be considered valuable by graduating seniors. This post-hoc analysis of a non-experimental study is fraught with danger, however, and should be approached with caution. While the t values and ω^2's can legitimately be considered random variables, there is random fluctuation which is hard to take into account.

Be that as it may, those items at the top of the list and therefore most discriminating among the two groups appear to be those relating to broad, abstract educational outcomes: e.g., Gave me new viewpoints or appreciations, Helped broaden my interests, and Motivated me to do my best. The least discriminating items seem to be those relating to specific teaching behaviors, e.g., Clear and understandable in explanations, Made good use of examples and illustrations, and Material presented in a well-organized fashion.

This is a curious result in that it is items of the latter type that consistently show up as most important for good teaching. But having the mechanics of good teaching per se is apparently not sufficient to have a course chosen as one of the three most valuable during an entire undergraduate career, nor is poor techniques sufficient to have a course chosen as one of the three least valuable. These data suggest that the most impressive courses are those providing new and fresh perspectives, which broaden and motivate the students. One could argue that the enthusiastic presentation of material (the second most discriminating item) requires something beyond just clear explanation of the basic concepts of a field. Furthermore, the item "Abstract ideas and theories were clearly interpreted" is higher on the list than "Clear and understandable in explanations" which also is possibly indicative of the instructor's success at going beyond the basic data.

One final note on this question. The least discriminating item, "Helpful to Individual Students," gives indirect evidence that the path to being considered a valuable course by some number of students is not necessarily through spending a lot of time with individual students.

Reference Notes

de Wolf, V. A. The 1974 College of Arts and Sciences survey of graduating seniors (EAC Report 271). Seattle: University of Washington, Educational Assessment Center, 1975.

References

Hays, W. L. Statistics for psychologists. New York: Holt, Rinehart and Winston, 1963.